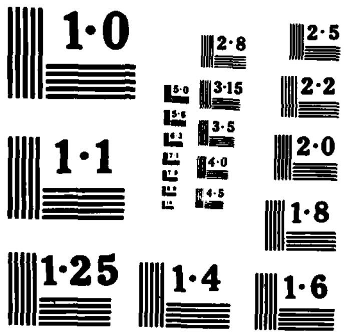
AD-A158 879 DEGENERATE MULTIVARIATE STATIONARY PROCESSES: BASICITY PAST AND FUTURE AN..(U) NORTH CAROLINA UNIV AT CHAPEL HILL CENTER FOR STOCHASTIC PROC.. A G MIAMEE ET AL. 1/1

UNCLASSIFIED MAY 85 TR-99 AFOSR-TR-85-0675 F/G 12/1 NL

END

FILMED

DTIC




NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

## ORT DOCUMENTATION PAGE

# AD-A158 879

1b. RESTRICTIVE MARKINGS

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release
Distribution Unlimited.

(2)

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
Technical Report No. 99

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 85-0675

6a. NAME OF PERFORMING ORGANIZATION
Center for Stochastic Processes

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NM

6c. ADDRESS (City, State and ZIP Code)
Dept. of Statistics
University of NOrth Carolina
Chapel Hill, NC 27514

7b. ADDRESS (City, State and ZIP Code)
Bldg. 410
Bolling AFB D.C. 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F49620-82-C-0009

8c. ADDRESS (City, State and ZIP Code)
Bolling AFB
Washington, DC 20332

10. SOURCE OF FUNDING NOS.

| PROGRAM ELEMENT NO. | PROJECT NO. | TASK NO. | WORK UNIT NO. |
|---|---|---|---|
| 61102F | 2304 | A5 | |

11. TITLE (Include Security Classification)
DEGENERATE MULTIVARIATE STATIONARY PROCESSES: BASICITY, PAST AND FUTURE, AND AUTOREGRESSIVE REPRESENTATION

12. PERSONAL AUTHOR(S)
A.G. Miamee and M. Pourahmadi

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT (Yr., Mo., Day) | 15. PAGE COUNT |
|---|---|---|---|
| Technical | FROM 9/84 TO 8/85 | May 1985 | 37 |

16. SUPPLEMENTARY NOTATION

| 17. COSATI CODES | | |
|---|---|---|
| FIELD | GROUP | SUB. GR. |
| | | |
| | | |

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
q-variate stationary processes, autoregressive and moving average representations, spectral density, mean convergence, angle between past and future

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Let $\{X_n\}$ be a not necessarily full rank multivariate weakly stationary stochastic process. It is shown that $\{X_n\}$ forms a generalized Schauder basis for the time domain of the process if and only if the angle between its past-present and future subspaces is positive. Then validity of the autoregressive representation of $\{X_n\}$, and that of its predictor, is considered and some characterization for these representations are given. Under the additional assumption that the range of the spectral density f of a degenerate process $\{X_n\}$ is constant some more concrete criteria for the validity of these representations are obtained.



20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☑ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

| 22a. NAME OF RESPONSIBLE INDIVIDUAL | 22b. TELEPHONE NUMBER (Include Area Code) | 22c. OFFICE SYMBOL |
|---|---|---|
| Brian W. Woodruff MAJ, USAF | (202)767-5028 | NM |

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

AFOSR-TR- 85-0675

# CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

DEGENERATE MULTIVARIATE STATIONARY PROCESSES:

BASICITY, PAST AND FUTURE, AND AUTOREGRESSIVE REPRESENTATION

by

A.G. Miamee

and

Mohsen Pourahmadi



Technical Report No. 99

May 1985

# DEGENERATE MULTIVARIATE STATIONARY PROCESSES: BASICITY, PAST AND FUTURE, AND AUTOREGRESSIVE REPRESENTATION



A.G. Miamee[1]
Center for Stochastic Processes
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

and

Mohsen Pourahmadi[2]
Department of Mathematical Sciences
Northern Illinois University
DeKalb, IL 60115

Abstract: Let $\{X_n\}$ be a not necessarily full rank multivariate weakly stationary stochastic process. It is shown that $\{X_n\}$ forms a generalized Schauder basis for the time domain of the process if and only if the angle between its past-present and future subspaces is positive. Then validity of the autoregressive representation of $\{X_n\}$, and that of its predictor, is considered and some characterization for these representations are given. Under the additional assumption that the range of the spectral density f of a degenerate process $\{X_n\}$ is constant some more concrete criteria for the validity of these representations are obtained.

AMS 1980 Subject Classification: Primary 62M10; Secondary 60G12

Keywords and Phrases: q-variate stationary processes, autoregressive and moving average representations, spectral density, mean convergence, angle between past and future.

[1]Research supported by AFOSR Contract No. F49620 82 C 0009. On leave from Isfahan University of Technology, Isfahan, Iran.

[2]Research supported by the NSF Grant MCS-8301240.

1. Introduction. Let $\{X_n\} = \{X_n;\ n = 0, \pm 1, \ldots\}$ be a q-variate weakly stationary stochastic process (WSSP) with spectral distribution (matrix) $F(\lambda)$, $-\pi < \lambda \leq \pi$.

weakly stationary stochastic (WSSP)

An important problem in prediction theory of ~~such~~ processes is to find conditions on the process, or equivalently on its spectral distribution F, so that the linear least square predictor of a future value of the process admits a mean-convergent series representation in terms of the past (observed) values of the process. This problem was solved by Wiener and Masani [18], and Masani ~~[6]~~ by imposing some boundedness conditions on f, the spectral density (matrix) function of the process.

Recently, using the notion of positivity of the angle between the past-present and the future subspaces of the process it was shown by Pourahmadi ~~[13, 14, 15]~~ that the series representation of the predictor is possible under some weaker conditions. This was made possible by using the idea of angle due to Helson and Szegö [5] (see also Miamee [10] for a multivariate extension of this). However these results hold under conditions which require the process to be of full rank. The main purpose of the ~~present study~~ this document is to consider the same problem, including their autoregressive representation, for the degenerate WSSP's. Additional keywords: moving average representation.

Now we explain the results of this paper .

It is well-known [7] that every purely nondeterministic WSSP has a one-sided moving average representation as

$$(1.1) \qquad X_n = \varepsilon_n + C_1\varepsilon_{n-1} + \ldots = \sum_{k=0}^{\infty} C_k \varepsilon_{n-k}, \quad (C_0 = I),$$

where $\{\varepsilon_n\}$ is the innovation process of $\{X_n\}$ and $\{C_k\}$ is a sequence of $q \times q$ constant matrices with

$$\sum_{k=0}^{\infty} \operatorname{tr} C_k C_k^* < \infty.$$

This moving average representation plays an extremely important role in prediction theory and statistical analysis of $\{X_n\}$. For example, from (1.1) one can obtain the $\nu$-step ahead predictor of $X_{n+\nu}$ (denoted by $\hat{X}_{n+\nu}$) based on $X_n, X_{n-1}, \ldots,$ as

$$\hat{X}_{n+\nu} = \sum_{k=\nu}^{\infty} C_k \varepsilon_{n+\nu-k}, \tag{1.2}$$

$$X_{n+\nu} - \hat{X}_{n+\nu} = \sum_{k=0}^{\nu-1} C_k \varepsilon_{n+\nu-k}. \tag{1.3}$$

Also the moving average representation (1.1) is used in studying the limiting distribution of certain statistics, such as the estimators of the autocovariance and the spectral density matrix, which are useful in the analysis of time series data collected from $\{X_n\}$.

To render the moving average representation (1.1) and the form of the best linear predictor fully satisfactory we should be able to express the innovation process $\varepsilon_n$ in terms of the past (observed) values of the process $X_n$ itself, so that the best linear predictor would be also expressed in terms of these observed values of the process.

<u>1.4 Definition.</u> We say that the moving average processes $X_n$ in (1.1) has an *autoregressive representation* if there exists a sequence $\{A_k\}$ of q×q matrices such that

$$\varepsilon_n = \sum_{k=0}^{\infty} A_k X_{n-k}, \tag{1.5}$$

where the infinite series is to converge in the mean.

It is useful to note that (1.1) can be viewed as a stochastic difference equation in $\{\varepsilon_n\}$ with $X_n$ as the input. The existence of the autoregressive

representation (1.5) assures that the difference equation (1.1) has a solution and therefore (1.5) can be viewed as a stochastic difference equation in $\{X_n\}$ with $\varepsilon_n$ as the input, i.e. the roles of $X_n$ and $\varepsilon_n$ can be reversed. Due to this reversal of roles, the term *invertibility* of the moving average (1.1) is sometimes used in the literature on time series, instead of the autoregressive representation.

Note that if $\{X_n\}$ has an autoregressive representation, then it follows from (1.5) that the one step ahead predictor of $\{X_n\}$ satisfies the equation

$$A_0 \hat{X}_{n+1} = \sum_{k=1}^{\infty} A_k X_{n+1-k}, \tag{1.6}$$

which can be solved for $\hat{X}_{n+1}$, expressing it (uniquely) in terms of the observed values of the process, provided that $A_0$ is invertible. However, as we shall see in section 4 the invertibility of $A_0$ is tied up with the invertibility of the prediction error matrix G. Thus the question of the rank of G, or equivalently the rank of the process $\{X_n\}$, enter the scene. Also, it follows from (1.5), upon formal substitution in (1.2), that the autoregressive representation of $\{X_n\}$ may entail an autoregressive representation of its linear least square predictor.

1.7 Definition. Let $\{X_n\}$ be as in (1.1) and $\nu \geq 1$ be a fixed integer. We say that the linear least square predictor of $\{X_n\}$ has an *autoregressive representation* if there exists a sequence of constant q×q matrices $\{E_{\nu k}\}_{k=0}^{\infty}$ such that

$$\hat{X}_{n+\nu} = \sum_{k=0}^{\infty} E_{\nu k} X_{n-k}, \tag{1.8}$$

where the infinite series is to converge in the square mean.

In the light of Definitions 1.4 and 1.7 it is natural to ask how the

autoregressive representation of $\{X_n\}$ and that of its linear least square predictor are related. In section 4 we show that these two representation are, indeed, equivalent regardless of the rank of the process $\{X_n\}$, and in fact $A_0$ can be taken to be I. This result shows the importance of the autoregressive representation problem of $\{X_n\}$ in prediction theory. To solve this problem one has to find conditions on the spectral distribution F, so that the infinite series in (1.5) converges in the mean, which is in turn equivalent to the convergence of

$$\sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \operatorname{tr} A_k \Gamma_{\ell-k} A_\ell^* < \infty.$$

Although our initial results in sections 3 and 4 have been worked out in the time domain, which is useful as to the application in time series is concerned, to get such spectral criteria for the autoregressive representation problem we have to move to the spectral domain.

After setting up the notations and preliminary results in section 2, we consider in section 3, a problem which is more general than the autoregressive representation, namely, to find conditions on F which enables one to write any Y in the time domain of $\{X_n\}$ as a unique series expansion in terms of $X_n$'s. The main result here is Theorem 3.8, which provides some characterization for this general property. An important consequence of this characterization and the other results in this section is the fact that if the angle between past and future is positive then the range of the spectral density $f(\lambda)$ is constant. This fact, together with a technique used by Miamee and Salehi [12]) (cf. also [11]) reduces our problem regarding a degenerate rank q-variate WSSP to the same problem for a corresponding full rank p-variate ($p \leq q$) WSSP.

In section 4 we prove the equivalence of the autoregressive representation

of $X_n$ and that of its predictor and we show that this happens if and only if the Fourier (Taylor) series of the reciprocal $\Omega^{-1}$ of the factor $\Omega$ of the generating function $\Phi = \Omega \sqrt{G}$ of the process, converges to it in the norm of $L^2(f)$ (This is Corollary 4.16). We should say here that *it seems there is no better spectral criterion for the autoregressive representation, without further restrictions on* $X_n$. This is important in view of Professor Masani's call to find a *good* necessary and sufficient condition for the validity of the autoregressive representation for $X_n$ or its predictor.

Of course, the necessary and sufficient condition just mentioned is not very useful because it is not expressed directly in terms of the spectral density. However, using this in conjunction with our other results, we give several concrete and useful sufficient conditions for the validity of the autoregressive representation. By a theorem of Matveev [9], the density of every purely non-deterministic process has constant rank and our work in section 4 is under the additional requirement that the range of f is constant, which is of course motivated by the results of section 3 mentioned above. It is also shown that analogues of the sufficient conditions due to Masani [6] and Pourahmadi [14,15] holds true in the degenerate rank as well.

We finally remark that in the presentation of this study a special attempt is made to work in the time domain as far as possible. Such a program is particularly useful for the purpose of applications. Also this, and especially working with the nonnormalized process $\{\varepsilon_n\}$, is helpful as it postpones the complications arising from the degeneracy of the rank of f, and results in a factorization of $f(\lambda)$ in the form $\Omega(\lambda)G\Omega(\lambda)^*$, which whows that the degeneracies of $f(\lambda)$ stem from a constant matrix, namely the prediction error matrix G. (For more on this see [7, Theorem 13.3].

2. Preliminaries. Let $(\Omega, G, P)$ be a probability space. $H = L_0^2(\Omega, G, P)$ denotes the Hilbert space of all complex-valued random variables on $\Omega$ with zero expectation and finite variance. The inner product in $H$ is given by

$$(x,y) = Ex\bar{y}, \quad x,y \in H.$$

Following [7], for $q \geq 1$, $H^q$ denotes the Cartesian product of $H$ with itself $q$ times, i.e. the set of all column vectors $X = (x_1, x_2, \ldots, x_q)^T$ with $x_i \in H$, $i = 1,2,\ldots,2$. $H^q$ is endowed with a Gramian structure: For $X$ and $Y$ in $H^q$ their *Gramian* is defined to be the $q\times q$ matrix $(X,Y) = [(x_i, y_j)]_{i,j=1}^q$. $H^q$ is a Hilbert space under the inner product $((X,Y)) = \text{trace } (X,Y) = \sum_{j=1}^q (x_j, y_j)$ and norm $||X|| = \sqrt{((X,X))}$ provided the linear combinations are formed with constant $q\times q$ matrices as coefficients.

For a $q\times q$ matrix $A = (a_{ij})$, $\text{tr } A = \sum_{j=1}^q a_{jj}$, $A^* = (\bar{a}_{ji})$ and $\det A$ stands for determinant of $A$. When $A$ is singular $A^\#$ denotes its Moore-Penrose generalized inverse. Functions are defined on $(-\pi,\pi]$ and we identify this interval with the unit circle in the complex plane in the natural way. Typical values of a function $f$ defined on $(-\pi,\pi]$ or on the unit circle will be denoted by $f(\lambda)$. $dm$ denotes the normalized Lebesgue measure on $(-\pi,\pi]$. For $1 \leq p \leq \infty$, $L^p(H^p)$ denotes the usual Lebesgue (Hardy) space of functions on the unit circle. $L^p_{q\times q}(H^p_{q\times q})$ denotes the space of all $q\times q$ matrix-valued functions whose entries are in $L^p(H^p)$.

Let $\{X_n, n = 0,\pm 1,\ldots\} \subset H^q$. It is said that $\{X_n\}$ is a q-variate weakly stationary stochastic process (WSSP), if the Gramian matrix $(X_m, X_n)$ depends only on $m-n$. It can be shown [7] that such a process has a *spectral representation* of the form

$$X_n = \int_{-\pi}^{\pi} e^{-in\lambda}\, dZ(\lambda),$$

where $Z(\cdot)$ is a countably additive orthogonally scattered $H^q$-valued measure. The $q\times q$ nonnegative matrix valued measure $F(\cdot) = (Z(\cdot),Z(\cdot))$ is called the *spectral distribution* of $\{X_n\}$. In case $F \ll dm$, we say that $\{X_n\}$ has the *spectral density* $f = F' = \frac{dF}{dm}$. The *spectral domain* corresponding to the spectral distribution F is denoted by $L^2(dF)$ and is defined by

$$L^2(dF) = \{\Psi;\ \Psi \text{ is a } q\times q \text{ matrix valued function with } \|\Psi\|_F^2 =$$

$$\int_{-\pi}^{\pi} \operatorname{tr}\Psi(\theta)\, dF(\theta)\, \Psi^*(\theta) < \infty\}.$$

It is well-known [7] that $L^2(dF)$ with inner product

$$((\Phi,\Psi))_F = \int_{-\pi}^{\pi} \operatorname{tr}\Phi\, dF\Psi^*$$

is a Hilbert space.

For each subset $\{\ldots\}$ of $H^q$, $\overline{sp}\{\ldots\}$ stands for the closed linear span of elements of $\{\ldots\}$ in the metric of $H^q$ and the following subspaces associated to our process $\{X_n\}$ are needed.

$$H(X) = \overline{sp}\{X_k;\ k = 0,\pm 1,\ldots\},$$

$$P_n(X) = \overline{sp}\{X_k;\ k \le n\},\ n = 0,\pm 1,\ldots,$$

$$F_n(X) = \overline{sp}\{X_k;\ k \ge n\},\ n = 0,\pm 1,\ldots,$$

$$P_{-\infty}(x) = \bigcap_{n=-\infty}^{+\infty} P_n(x),$$

$$M_n(X) = \overline{sp}\{X_k;\ k \ne n\},\ n = 0,\pm 1,\ldots,$$

$$M_{-\infty}(x) = \bigcap_{n=-\infty}^{+\infty} M_n(X).$$

The space $H(X)$ is referred to as the *time domain* of the process $\{X_n\}$. It is well-known [7] that the correspondence

$$T: \Psi \to \int_{-\pi}^{\pi} \Psi(\lambda)\, dZ(\lambda)$$

is an isometric isomorphism from $L^2(dF)$ onto $H(X)$. T is called the *Kolmogorov isomorphism* between the spectral and time domain, and plays an important role in finding analytical conditions, in terms of F, for the following important geometrical (regularity) properties of the WSSP $\{X_n\}$.

2.1 Definition: Let $\{X_n\}$ be a q-variate WSSP

a) $\{X_n\}$ is said to be *purely nondeterministic* (regular) if

$$P_{-\infty}(X) = \{0\}.$$

b) $\{X_n\}$ is said to be *minimal* if for some n

$$M_n(X) \neq H(X)$$

c) $\{X_n\}$ is said to be *$J_0$-regular if*

$$M_{-\infty}(X) = \{0\}.$$

d) It is said that the past-present and the future subspaces of $\{X_n\}$ are at *positive angle* if

$$\rho(X) = \rho(F) < 1,$$

where

$$\rho(X) = \rho(F) = \sup\{|(Y,Z)|: Y \in P_0(X),\ Z \in F_1(X) \text{ and } \|Y\| = 1,\ \|Z\| = 1\}.$$

Let $\{X_n\}$ be a purely nondeterministic WSSP. The best linear predictor of $X_{n+\nu}$, $\nu \geq 1$, based on $X_n$, $X_{n-1}, \ldots$ is given by

$$\hat{X}_{n+\nu} = (X_{n+\nu} | P_n(X)),$$

where the latter denotes the orthogonal projection of $X_{n+\nu}$ on the subspace $P_n(X)$ of $H(X)$. For such a process we define a new process $\{\varepsilon_n\}$ by

$$\varepsilon_n = X_n - (X_n | P_{n-1}(X)), \; n = 0, \pm 1, \ldots,$$

and it is called the *innovation process* of $\{X_n\}$. It is known that $\{\varepsilon_n\}$ satisfies $(\varepsilon_m, \varepsilon_n) = \delta_{m,n} G$, and $G$ is called the prediction error matrix of lag 1. A WSSP $\{X_n\}$ is said to be of *full rank* if its matrix $G$ is invertible (full rank). Otherwise, the process $\{X_n\}$ is of *degenerate rank.*

It follows from (1.1) that the spectral density $f(\lambda)$ of a purely nondeterministic process admits a factorization of the form

$$f(\lambda) = \Omega(\lambda) G \Omega(\lambda)^* = \Omega(\lambda) G^{1/2} \, [\Omega(\lambda) G^{1/2}]^*,$$

where

$$\Omega(\lambda) = \sum_{k=0}^{\infty} C_k e^{ik\lambda},$$

is in $H^2_{q \times q}$. It is shown in [7] that $\Omega(\cdot)$ is an almost everywhere invertible function. Thus, it follows that when $\{X_n\}$ is not of full rank then $f(\cdot)$ is not invertible.

## 3. Basicity and positivity of the angle.

In this section we will define the idea of generalized Schauder basis for a set of vectors and also give the definition of the angle between past-present and future for a q-variate WSSP, and then we give several criteria for a WSSP $X_n$ to form a Schauder basis for its time domain $H(X)$. We will also get several other results which are essential in dealing with the problem of autoregressive representation for degenerate rank multivariate processes in the next section.

Because of the imporatnce of the mean-square convergences in different areas of applications (particularly time series analysis) it seems that the idea of $\{X_n\}$ forming a generalized Schauder basis is more appropriate than the weaker requirement of $\{X_n\}$ forming a conditional basis as studied by Rozanov [17, pp. 104-108].

To get a feeling as to how the question of uniqueness of the representation of elements of $H(X)$ in terms of a sequence will arise and should be settled in the non full rank case we start this section with an example:

3.1 Example. Let $\{e_n\}$ be a univariate white noise process, i.e. $Ee_m\bar{e}_n = \delta_{m,n}$, and let $\{X_n\}$ be the bivariate process defined by

$$X_n = \begin{pmatrix} e_n \\ 0 \end{pmatrix}, \quad n = 0, \pm 1, \pm 2, \ldots .$$

For a fixed k, consider the element $Y = \begin{pmatrix} e_k \\ e_k \end{pmatrix} \in H(X)$. Note that for this element we have several different representations in terms of $\{X_n\}$, viz.

$$Y = \begin{pmatrix} e_k \\ e_k \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} X_k = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} X_k = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} X_k ,$$

and yet

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} .$$

i.e. there is no unique representation for Y in terms of $X_n$'s.

In view of this simple example and the need for uniqueness of the linear representation in terms of $X_n$'s in practical problems, it is important to find conditions on $\{X_n\}$ so that every element of H(X) has a linear representation in terms of $\{X_n\}$ which is unique in some sense.

Next we define two kinds of uniqueness for linear representation of elements of H(X). Throughout this section $\{A_k\}$ denotes an arbitrary sequence of $q \times q$ matrices, and it is understood that the infinite series $\sum_{n=-\infty}^{\infty} A_n X_n$ converges in the norm of $H^q$ or in the square mean.

3.2 Definition. Let $\{X_n, n=0,\pm 1,\pm 2,\ldots\}$ be a q-variate process in $H^q$.

a) $\{X_n\}$ is said to be a *Schauder basis* for H(X) if every $Y \in H(X)$ has a unique representation

$$Y = \sum_{n=-\infty}^{\infty} A_n X_n ,$$

in the sense that if Y has another representation

$$Y = \sum_{n=-\infty}^{\infty} A_n' X_n ,$$

then,

$$A_n = A_n' , \text{ for all } n.$$

b) It is said that $\{X_n\}$ is a *generalized Schauder basis* for H(X) if every

$Y \in H(X)$ can be represented as

$$Y = \sum_{n=-\infty}^{\infty} A_n X_n,$$

and furthermore this representation is unique in the sense that if $Y$ has another representation as

$$Y = \sum_{n=-\infty}^{\infty} A'_n X_n,$$

then,

$$A_n X_n = A'_n X_n, \text{ for all } n.$$

3.3 Remarks.

a) It should be noted that the notion of Schauder basis for $H(X)$ defined in 3.2(a) is different from that defined in the literature on classical Banach spaces. Since here $H(X)$ is a linear space over the ring of $q \times q$ matrices instead of the field of scalars. In this setting it is possible to develop a theory of Schauder basis for $H(X)$ which is similar, but technically different from that for classical Banach spaces.

b) For the classical Banach spaces the notions of Schauder and generalized Schauder basis are equivalent. This is not the case for $H(X)$ as Example 3.1 shows. In this example $\{X_n\}$ is a generalized Schauder basis but not a Schauder basis.

Since a generalized Schauder basis for $H(X)$ is not necessarily a Schauder basis, it is of interest to impose conditions on $\{X_n\}$ or $F$ so that the two notions become equivalent. When $\{X_n\}$ is a q-variate WSSP we have:

3.4 Theorem. Let $\{X_n\}$ be a q-variate WSSP with the spectral distirbution $F$ and let $\Gamma_0 = (X_n, X_n) = \int_{-\pi}^{\pi} dF(\lambda)$. Then the following conditions are equivalent:

a) $\{X_n\}$ is a generalized Schauder basis for $H(X)$ and $\Gamma_0$ is invertible.

b) $\{X_n\}$ is a Schauder basis for $H(X)$.

Proof. a $\Rightarrow$ b.

*Let* $Y \in H(X)$ *have two representations, viz.*

$$Y = \sum_{n=-\infty}^{\infty} A_n X_n = \sum_{n=-\infty}^{\infty} A_n' X_n ,$$

*since* $\{X_n\}$ *is a generalized Schauder basis for* $H(X)$ *we have*

$$A_n X_n = A_n' X_n , \text{ for all } n.$$

*Thus*

$$A_n \Gamma_0 = A_n (X_n, X_n) = A_n' (X_n, X_n) = A_n' \Gamma_0, \quad \text{for all } n.$$

*But, since* $\Gamma_0$ *is invertible we get*

$$A_n = A_n', \text{ for all } n,$$

*i.e.* $\{X_n\}$ *is a Schauder basis for* $H(X)$.

b $\Rightarrow$ a.

*Suppose* $\{X_n\}$ *is a Schauder basis for* $H(X)$ *and* $\Gamma_0$ *is not invertible. Then there exists a non-zero vector* $\alpha = (\alpha_1, \ldots, \alpha_q)$ *such that*

$$\sum_{i=1}^{q} \alpha_i X_{n,i} = 0,$$

*This implies that for* $0 = Y \in H(X)$ *we have*

$$Y = 0 = \begin{pmatrix} \alpha_1 & \cdots & \alpha_q \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} X_{n,1} \\ \vdots \\ X_{n,q} \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} X_{n,1} \\ \vdots \\ X_{n,q} \end{pmatrix} ,$$

*and yet*
$$\begin{pmatrix} \alpha_1 \cdots \alpha_q \\ 0 \cdots 0 \\ \vdots \quad \vdots \\ 0 \cdots 0 \end{pmatrix} \neq \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

*which contradicts the assumption that* $\{X_n\}$ *is a Schauder basis for* $H(X)$. Q.E.D.

3.5 Remark. Since $\Gamma_0 \geq G$, it follows that when $\{X_n\}$ is of full rank, then the two notions of bases are equivalent.

The next theorem which provides a necessary and sufficient condition for a q-variate process $\{X_n\}$ to be a generalized Schauder basis for $H(X)$ is essential in the rest of this work. This theorem is a generalization of a well-known theorm of Nikolskii, cf. [4, p. 103], to the setting of $H(X)$. Although the main steps of its proof are the same as those in the classical setting, the details are different as the $A_k$'s here are q×q matrices instead of being complex scalars. Since the proof is lengthy we have relegated it to an appendix.

3.6 Theorem. A q-variate process $\{X_n\}$ is a generalized Schauder basis for $H(X)$, if and only if there exists a positive real number M such that

$$\left\|\sum_{n=k}^{\ell} A_n X_n\right\| \leq M \left\|\sum_{n=-\infty}^{\infty} A_n X_n\right\|,$$

for any $\ell \geq k$ and all q×q matrices $A_n$, $n = 0, \pm 1, \ldots$, with $\sum_{n=-\infty}^{\infty} A_n X_n \in H^q$.

In general, it is hard to verify the condition of Theorem 3.6. However, when $\{X_n\}$ is a WSSP, then as it turns out this condition is equivalent to the geometrical condition that the past-present and future subspaces of $\{X_n\}$ being at positive angle. This is summarized in the next lemma whose proof, being exactly the same as in the univariate case (cf. [3, pp. 129-130]), is omitted.

3.7 Lemma. Let $\{X_n\}$ be a q-variate WSSP. Then the following are equivalent:

a) $\rho(X)<1$.

b) There exists a positive number M such that

$$||\sum_{n=k}^{\ell} A_n X_n|| \leq M||\sum_{n=-\infty}^{\infty} A_n X_n||,$$

for all $\ell \geq k$ and all q×q matrices $A_n$, $n = 0,\pm 1,\pm 2, \ldots$, with $\sum_{n=-\infty}^{\infty} A_n X_n \in H^q$.

Next, by combining Theorem 3.6 and Lemma 3.7 we get the following important result.

3.8 Theorem. Let $\{X_n\}$ be a q-variate WSSP. Then the following are equivalent:

a) $\rho(X)<1$.

b) $\{X_n\}$ is a generalized Schauder basis for H(X).

In view of Theorem 3.8 it is important to characterize WSSP's for which $\rho(X)<1$, i.e. to find spectral characterization for this useful geometrical property. When $\{X_n\}$ is a q-variate WSSP of full rank such characterization is given in [10,14,16]. It is important to note, however, that the techniques used in these papers do not work when $\{X_n\}$ is not of full rank.

Here we use a different method which is based on exploiting the characterization of $\rho(X)<1$ given in Theorem 3.8 and the uniqueness of the representation of elements of H(X) when $\{X_n\}$ is a generalized Schauder basis. As a result of this we show that when $\{X_n\}$ is a generalized Schauder basis for H(X), then $\{X_n\}$ is $J_0$-regular.

3.9 Theorem. Let $\{X_n\}$ be a q-variate WSSP with $\rho(X)<1$. Then $\{X_n\}$ is $J_0$-regular.

Proof. *Since $\rho(X)<1$, it follows from Theorem 3.8 that $\{X_n\}$ is a generalized*

*Schauder basis for* $H(X)$. *To show that* $\{X_n\}$ *is* $J_0$*-regular, let*

$$Y \in M_{-\infty}(X) = \bigcap_{n=-\infty}^{\infty} M_n(X). \quad \textit{Thus}$$

$$Y \in M_n(X), \quad \textit{for all } n.$$

*Now, as in the proof of the "if" part of Theorem 3.6 one can show that* $\{X_k;\ k \neq n\}$ *is a generalized Schauder basis for* $M_n(X)$. *Hence, there exists a series representation for* $Y$;

$$Y = \sum_{k \neq n} A_{k,n} X_k, \quad \textit{for all } n.$$

*From this and the uniqueness of such representations as defined in* 3.2 (b) *we get*

$$A_{k,n} X_n = 0, \quad \textit{for all } n,k,$$

*which implies that* $Y = 0$. Q.E.D.

As an immediate consequence of Theorem 3.9 we see that if $\rho(X)<1$, then the process $\{X_n\}$ is minimal, and purely nondeterministic. The following Corollary is very crucial for our purposes.

3.10 Corollary. Let $\{X_n\}$ be a q-variate WSSP with the spectral distribution F. If $\rho(X)<1$, then F has the following properties:

a) F is absolutely continuous with respect to the Lebesgue measure on $(-\pi,\pi]$.

b) $R(f)$ = constant supspace a.e., where f is the density of the process and $R(f)$ denotes the range of f when $f(\cdot)$ is viewed as an operator from $\mathbb{C}^q$ into $\mathbb{C}^q$.

c) $f^{\#} \in L^1_{q\times q}$, where $f^{\#}$ denotes the Moore-Penrose generalized inverse of the matrix f.

Proof. Since $\rho_X < 1$, *by Theorem 3.9 our WSSP* $\{X_n\}$ *is* $J_0$*-regular and it is shown by Makagon and Weron* [5, Theorem 5.3] *that* $\{X_n\}$ *is* $J_0$*-regular, if and only if* F *satisfies the conditions* a,b, *and* c.

Corollary 3.10 gives some very useful necessary conditions for $\rho(X)$ to be strictly less than one, the most important of these is the condition concerning the constancy of the range of f as the subsequent argument will reveal. Thus in the following we work under the following natural assumption:

3.11 Assumption.

(i) $F << dm$,

(ii) $R(f)$ = constant a.e.(dm),

(iii) $f^{\#} \in L^1_{q\times q}$.

Under this assumption our main problem is reduced to characterizing WSSP's for which $\rho(X)<1$. Since $R(f)$ is a constant subspace of $\mathbb{C}^q$, we let $R$ denote this subspace and $p(0 < p \leq q)$ its dimension. It follows from the proof of Theorem 3.1 in [12] that there exists a $q\times q$ constant unitary matrix U such that

$$Uf(\cdot)U^* = \left(\begin{array}{c|c} g(\cdot) & 0 \\ \hline 0 & 0 \end{array}\right), \tag{3.12}$$

where $g(\cdot)$ is a $p\times p$ matrix-valued function on $(-\pi,\pi]$. It is easy to check that g is the spectral density of a p-variate purely nondeterministic full rank WSSP, say, $\{Y_n\}$. This matrix U and the WSSP $\{Y_n\}$ play important roles in what follows and their relationship with f and $\{X_n\}$ is prescribed by (3.12), throughout this paper. An important consequence of this relationship is the following.

<u>3.13 Theorem</u>. Let $\{X_n\}$ be a q-variate (not necessarily full rank) WSSP whose spectral distribution F satisfies (3.11). Then, with notation as above, we have

$$\rho(X) < 1, \quad \text{if and only if } \rho(Y) < 1.$$

<u>Proof</u>. *It follows from* 3.12 *that*

$$U X_n = \begin{bmatrix} Y_n \\ \hline 0 \end{bmatrix},$$

*thus for each* $Z \in H(Y)$ *we have*

$$\eta = U^* \begin{bmatrix} Z \\ \hline 0 \end{bmatrix} \in H(X).$$

*Assume that* $\rho(X) < 1$. *Then by Theorem 3.8,* $\eta$ *has a representation,*

$$\eta = \sum_{n=-\infty}^{\infty} A_n X_n.$$

*Therefore,*

$$U\eta = \sum_{n=-\infty}^{\infty} U A_n U^* U X_n,$$

*which implies that, with* $C_n = UA_nU^*$, *we have*

$$\begin{bmatrix} Z \\ \hline 0 \end{bmatrix} = \sum_{n=-\infty}^{\infty} C_n \begin{bmatrix} Y_n \\ \hline 0 \end{bmatrix},$$

*or*

$$Z = \sum_{n=-\infty}^{\infty} D_n Y_n,$$

*where* $D_n$ *is the* $p \times p$ *leading principal minor of* $C_n$. *By using the uniqueness of the representation of* $\eta$ *one can show that this representation of* $Z$ *is unique in the sense of Definition 2.2(b). Thus* $\{Y_n\}$ *is a generalized Schauder basis*

*for H(Y) and therefore by Theorem 3.8 we get that*

$$\rho(X)<1.$$

*The proof of the other direction is similar.* Q.E.D.

Theorem 3.13 reduces the problem of characterizing non-full rank processes with $\rho(X)<1$ to that for full rank processes with smaller dimension. In view of this one can invoke the known results in the full rank case and state the appropriate conditions in terms of g, the spectral density matrix of the corresponding full rank process $\{Y_n\}$. We should note that the statement of the next theorem is not correct if one replaces g by f, as Example 3.1 provides a counterexample to this effect. Proof of the next theorem is immediate from 3.13 and the results of [10].

3.14 Theorem. Let $\{X_n\}$ be as in Theorem 3.13, then the following are equivalent:

a) $\rho(X)<1$.

b) $L^2(g) \subset L^1_{p\times p}$ and the Fourier series of any function $\psi$ in $L^2(g)$ converges to $\psi$ in the norm of $L^2(g)$.

Although Theorem 3.14 provides necessary and sufficient conditions for $\rho(X)<1$ in terms of the spectral domain $L^2(g)$ of $\{Y_n\}$, it does not provide any concrete conditions involving the entries of the spectral density. In the following we review some known results which provide more tangible conditions for $\rho(X)<1$. However, in light of Theorem 3.11, we state all these results for a full rank WSSP.

In the univariate case a complete characterization of WSSP with $\rho(X)<1$ is

given by Helson and Szegö [3].

3.15 Theorem. Let $\{X_n\}$ be a univariate WSSP with density f. Then $\rho(X)<1$, if and only if

$$f = e^{u+\tilde{v}} , \tag{3.16}$$

where u and v are bounded real-valued functions with $||v||_\infty < \frac{\pi}{2}$, and $\tilde{v}$ denotes the harmonic conjugate of the function V.

For q-variate processes one can find such a characterization provided that f has some special properties. For a q×q matrix-valued function, denote the smallest and largest eigenvalues of $f(\lambda)$ by $f_1(\lambda)$ and $f_q(\lambda)$, respectively. With these notations we have [14, Theorem 5.3].

3.17 Theorem. Let $\{X_n\}$ be a q-variate purely nondeterministic full rank WSSP with a spectral density f satisfying

$$\frac{f_q}{f_1} \in L^\infty . \tag{3.18}$$

Then $\rho(X)<1$, if and only if $f_q$ satisfies (3.16).

This theorem provides a characterization for $\rho(X)<1$ in terms of the largest eigenvalue of $f(\cdot)$ which in general might be hard to apply. The next lemma provides a necessary condition for $\rho(X)<1$ in terms of the diagonal entries of the density matrix f. Proof of this lemma is immediate from the definitions of $\rho(X)$ and the Kolmogorov's isomorphism.

3.19 Lemma. Let $\{X_n\}$ be a purely nondeterministic WSSP with the spectral density f. If $\rho(X)<1$, then for each $j = 1,2,\ldots,q$, $f_{jj}$ is the diagonal element of f which satisfies (3.16).

It is certainly of interest to show that the condition of Lemma 3.19 is also sufficient for $\rho(X)<1$. However, this is not true in general. It turns out that such a condition is necessary and sufficient for $\rho(X)<1$, when f is nearly diagonal [1]. Let $f = (f_{ij})_{i,j=1}^{q}$ and $f^{-1} = (f^{ij})_{i,j=1}^{q}$ be its inverse, with this notation we have the following definition due to Bloom [1].

3.20 Definition. An almost everywhere invertible matrix-valued function f is said to be *nearly diagonal*, if

$$||f_{ii}f^{ii}||_{\infty} < \infty \;, \; i = 1,2,\ldots,q.$$

It is easy to see that every diagonal density f is nearly diagonal, but the converse is not true. The next lemma provides a large class of nearly diagonal matrices which are not necessarily diagonal.

3.21 Lemma. If f satisfies (3.18), then f is nearly diagonal.

The following important theorem provides a necessary and sufficient condition for a q-variate WSSP $\{X_n\}$ with a nearly diagonal density to have the property $\rho(X)<1$.

3.22 Theorem. Let $\{X_n\}$ be a q-variate purely nondeterministic full rank WSSP with a nearly diagonal spectral density matrix f. Then $\rho(X)<1$, if and only if for every j, $1 \leq$ j $\leq q$, $f_{jj}$ satisfies (3.16).

Proof. It is immediate from Propositions 4.2, 4.5 [1].

## 4. AUTOREGRESSIVE REPRESENTATION OF WSSP's.

In this section we establish the equivalence of the autoregressive representation of a WSSP $\{X_n\}$ and that of its linear least squares predictor based on the past. As a consequence of this we obtain a spectral necessary and sufficient condition for the latter. This spectral characterization is used to provide more concrete and sufficient conditions, in terms of f, for the autoregressive representation of $\hat{X}_{n+\nu}$.

We have from the moving average representation (1.1) of $\{X_n\}$, and that $\{\varepsilon_n\}$ is its nonnormalized innovation process:

(4.1) $$(\varepsilon_m, \varepsilon_n) = \delta_{m,n} G, \quad \text{for integers } m,n,$$

(4.2) $$(X_m, \varepsilon_n) = 0, \quad m < n.$$

For the time being we assume that $\{X_n\}$ has a mean-convergent autoregressive representation as in (1.5) for a sequence $\{A_k\}$. First we attempt to express these $A_k$'s in terms of the $C_k$'s in (1.1). We do this in order to show in a natural way the importance of the rank of G in this determination. However, as we shall see this time domain procedure does not provide a satisfactory solution to our problem when $\{X_n\}$ is not of full rank. Therefore, when $\{X_n\}$ is of degenerate rank we appeal to a spectral domain argument and resolve the problem of determining the $A_k$'s in its full generality.

From (4.1) and (4.2) for any $\ell \geq 0$ and any integer n, we have

$$\delta_{0,\ell} G = (\varepsilon_n, \varepsilon_{n-\ell}) = \sum_{k=0}^{\infty} A_k (X_{n-k}, \varepsilon_{n-\ell}) = \sum_{k=0}^{\ell} A_k (X_{n-k}, \varepsilon_{n-\ell}),$$

which combined with

$$(X_{n-k}, \varepsilon_{n-\ell}) = C_{\ell-k} G, \quad k, \ell \geq 0,$$

gives

$$(4.3) \qquad \sum_{k=0}^{\ell} A_k C_{\ell-k} G = \delta_{0,\ell} \ , \ \text{for all } \ell \geq 0,$$

or equivalently (since $C_0 = I$),

$$(4.4) \qquad \begin{cases} A_0 G = G, \\ \sum_{k=0}^{\ell} A_k C_{\ell-k} G = 0 \quad (\text{or } A\ G = -\sum_{k=0}^{\ell-1} A_k C_{\ell-k} G), \ \ell > 0. \end{cases}$$

This shows the relationship among the matrices $A_k$'s, $C_k$'s and G, when $\{X_n\}$ has a mean-convergent autoregressive representation. In particular, it follows that when G is not of full rank, then $A_0, A_1, \ldots$ can not (necessarily) be found uniquely in terms of $C_k$'s.

To reveal more fully the role of G in determining the $A_k$'s, we need to introduce some notation. Corresponding to the moving average representation (1.1) of $\{X_n\}$ we consider the matrix-valued function $\Omega$ on $(-\pi,\pi]$ given by

$$(4.5) \qquad \Omega(\lambda) = \sum_{k=0}^{\infty} C_k e^{ik\lambda} \ ,$$

it is known [7, Theorem 13.3] that for almost every $\lambda$ in $[-\pi,\pi]$ this matrix function $\Omega$ is invertible and that the spectral density matrix $f(\lambda)$, then admits the factorization

$$(4.6) \qquad f(\lambda) = \Omega(\lambda) G \Omega(\lambda)^* = [\Omega(\lambda) G^{1/2}] \ [\Omega(\lambda) G^{1/2}]^*,$$

where $\Phi = \Omega G^{1/2}$ is referred to as the generating function of the WSSP $\{X_n\}$.

Let the matrix-valued function $\Psi(\lambda)$ denote the isomorph of $\varepsilon_0$ in $L^2(f)$, then it follows from (1.5) that

$$\Psi(\lambda) = \sum_{k=0}^{\infty} A_k e^{ik\lambda}. \tag{4.7}$$

In light of (4.3) or (4.4) it is easy to check that the functions $\Omega, \Phi$ and $\Psi$ satisfy the equation

$$\Psi \; \Phi \sqrt{G} = \Psi \; \Omega \; G = G. \tag{4.8}$$

This shows that when G is not of full rank then one may not be able to express $\Psi$, the isomorph of $\varepsilon_0$, in terms of $\Phi$ or $\Omega$. However, when the process $\{X_n\}$ is of full rank or when G is invertible then from (4.4) and (4.8) one obtains explicit formulas expressing $A_k$'s in terms of the $C_k$'s and what is more important or even surprising is that from (4.8) we get $\Psi = \sqrt{G}\;\Phi^{-1} = \Omega^{-1}$.

The previous time domain argument along with the assumption of existence of mean-convergent autoregressive representations of $\{X_n\}$ was used to show the shortcoming of the commonly used time domain approach in handling problems related to degenerate rank WSSP's.

Next, we use a spectral domain argument to show that for any purely nondeterministic WSSP, $\Psi$, the isomorph of $\varepsilon_0$ in $L^2(f)$ is given by $\Psi = \Omega^{-1}$.

Let $\Psi$ denote the isomorph of $\varepsilon_0$, then we have from (1.1) and the Kolmogovov's isomorphism that

$$Ie^{-in\lambda} = (e^{-in\lambda} + C_1 e^{-i(n-1)\lambda} + \dots)\Psi(\lambda) = e^{-in\lambda}\;\Omega(\lambda)\Psi(\lambda),$$

or $\Omega\Psi = I$. But, since $\Omega$ is invertible a.e., it follows that

$$\Psi\Omega = I = \Omega\Psi. \tag{4.9}$$

Thus we have

4.10 Lemma. Let $\{X_n\}$ be a purely nondeterministic q-variate WSSP with the spectral factorization (4.6). Then the isomorph of $\varepsilon_n$ in $L^2(f)$ is given by

$e^{-in\lambda}\Omega^{-1}(\lambda)$.

Another important consequence of (4.9) is the following set of identities which are crucial in the proof of Theorem 4.12.

$$(4.11)\qquad \begin{cases} A_0 = I \\ \sum_{k=0}^{\ell} A_k C_{\ell-k} = \sum_{k=0}^{\ell} C_{\ell-k} A_k = 0, \quad \ell > 0. \end{cases}$$

In the next theorem we establish the equivalence of the mean-convergent autoregressive representations of $\{X_n\}$ and that of its linear least squares predictor $\hat{X}_{n+\nu}$, $\nu \geq 1$. We note that the method of proof of this theorem is similar to that used by Bloomfield [2, Theorem 1]. But his method does not generalize to the case of q-variate degenerate rank WSSP's, because from (4.4) one can not conclude that $\sum_{k=0}^{\ell} C_{\ell-k} A_k = 0$, for $\ell > 0$, a fact which plays a crucial role in the proof of the theorem, cf. (4.15).

4.12 Theorem. Let $\{X_n\}$ be a q-variate purely nondeterministic (not necessarily full rank) WSSP with a one-sided moving average representation as in (1.1). Then the following are equivalent:

a) $\{X_n\}$ has a mean-convergent autoregressive representation.

b) For $\nu \geq 1$, we have

$$\hat{X}_{n+\nu} = \sum_{k=0}^{\infty} E_{\nu k} X_{n-k},$$

where $E_{\nu k} = -\sum_{j=1}^{\nu} C_{\nu-j} A_{j+k}$, $k = 0,1,2,\ldots$, and the infinite series defining $\hat{X}_{n+\nu}$ is convergent in the square mean.

Proof. (b) => (a) *is trivial in view of* (1.3).

*To show that* (a) => (b), *from* (1.3) *we get*

$$X_{n+\nu} - \hat{X}_{n+\nu} = \sum_{k=0}^{\nu-1} C_k \varepsilon_{n+\nu-k} = \sum_{j=1}^{\nu} C_{\nu-j} \varepsilon_{n+j}. \tag{4.13}$$

*Now from* (a) *or* (1.5) *we have*

$$\sum_{j=1}^{\nu} C_{\nu-j} \varepsilon_{n+j} = \sum_{j=1}^{\nu} C_{\nu-j} \sum_{k=0}^{\infty} A_k X_{n+j-k} = \sum_{j=1}^{\nu} C_{\nu-j} \{ \sum_{k=0}^{j-1} A_k X_{n+j-k} + \sum_{k=j}^{\infty} A_k X_{n+j-k} \}$$

$$= \sum_{j=1}^{\nu} C_{\nu-j} \sum_{k=0}^{j-1} A_k X_{n+j-k} + \sum_{j=1}^{\nu} C_{\nu-j} \sum_{k=0}^{\infty} A_{j+k} X_{n-k}$$

$$= \sum_{k=1}^{\nu} ( \sum_{j=k}^{\nu} C_{\nu-j} A_{j-k} ) X_{n+k} + \sum_{k=0}^{\infty} ( \sum_{j=1}^{\nu} C_{\nu-j} A_{j+k} ) X_{n-k}.$$

*Also, from* (4.11) *we get*

$$\sum_{j=k}^{\nu} C_{\nu-j} A_{j-k} = \sum_{s=0}^{\nu-k} C_{\nu-k-s} A_s = \begin{cases} C_0 A_0 = I, & \nu = k, \\ 0, & \nu > k. \end{cases} \tag{4.15}$$

*By combining* (4.13), (4.14) *and* (4.15) *we conclude that*

$$X_{n+\nu} - \hat{X}_{n+\nu} = X_{n+\nu} + \sum_{k=0}^{\infty} ( \sum_{j=1}^{\nu} C_{\nu-j} A_{j+k} ) X_{n-k},$$

*or*

$$\hat{X}_{n+\nu} = \sum_{k=0}^{\infty} E_{\nu k} X_{n-k}.$$

Q.E.D.

It is shown in Lemma 4.10 that the function $e^{-in\lambda}\Omega^{-1}(\lambda)$ is the isomorph of $\varepsilon_n$ in $L^2(F)$. Thus, it follows from the isomorphism between the time and spectral domains that $\{X_n\}$ has a mean-convergent autoregressive representation in $H(X)$, if and only if the Fourier series of $\Omega^{-1}$ converges to $\Omega^{-1}$ in the norm

of $L^2(f)$. This observation combined with Theorem 4.12 gives the following important characterization of processes $\{X_n\}$ which admit mean-convergent autoregressive representation.

4.16 Corollary. Let $\{X_n\}$ be as in Theorem 4.12. Then the following are equivalent:

a) $\{X_n\}$ has a mean-convergent autoregressive representation.

b) For $\nu \geq 1$, $\hat{X}_{n+\nu}$ has a mean-convergent autoregressive representation.

c) The Fourier series of $\Omega^{-1}$ converges to $\Omega^{-1}$ in the norm of $L^2(f)$.

4.17 Remark. It should be noted that for any density f with factorization as in (4.6), $\Omega^{-1} \varepsilon L^2(f)$. However it is not necessarily true that $\Omega^{-1} \varepsilon L^1_{q\times q}$. Thus, one may not be able to define the Fourier coefficients (and series) of the function $\Omega^{-1}$. One possible way of circumventing such difficulties is to work with the Taylor coefficients (and series) of $\Omega^{-1}$. This is possible since $\Omega^{-1}(z)$, $|z| < 1$, is an analytic function and therefore has a Taylor expansion. In this paper, however, we do not pursue this approach. Instead, by using some of the results of Section 3 we impose appropriate restrictions on f so that difficulties of the above type can not occur, as the following theorem shows.

4.18 Theorem. Let $\{X_n\}$ be as in Theorem 4.12 with $\rho(X)<1$. Then,

a) $\{X_n\}$ has a mean-convergent autoregressive representation.

b) For $\nu \geq 1$, $\hat{X}_{n+\nu}$ has a mean-convergent autoregressive representation.

Proof of this theorem is immediate from Theorem 3.8. Note that, since $\{X_n\}$ forms a generalized Schauder basis for $H(X)$ the autoregressive representation of $\{X_n\}$ or $\hat{X}_{n+\nu}$ is also unique (in the sense of Definition 3.2(b)). An

alternative proof of Theorem 4.18 can be obtained via Corollary 4.16 and Theorem 3.14 and observing (from (3.12)) that

$$\left[\begin{array}{c|c} g & 0 \\ \hline 0 & 0 \end{array}\right] = UfU^* = U\Omega U^*UGU^*U\Omega^*U^* = \left[\begin{array}{c|c} \Omega_{11} & \Omega_{12} \\ \hline \Omega_{21} & \Omega_{22} \end{array}\right]\left[\begin{array}{c|c} G_1 & 0 \\ \hline 0 & 0 \end{array}\right]\left[\begin{array}{c|c} \Omega_{11} & \Omega_{12} \\ \hline \Omega_{21} & \Omega_{22} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \Omega_{11}G_1\Omega_{11}^* & 0 \\ \hline 0 & 0 \end{array}\right],$$

where $G_1$ is the one-step prediction error matrix of $\{Y_n\}$ introduced in Section 3. It follows from (4.19) that the Fourier (Taylor) series of $\Omega^{-1}$ converges to $\Omega^{-1}$ in the norm of $L^2(f)$, if and only if the Fourier series of $\Omega_{11}^{-1}$ converges to $\Omega_{11}^{-1}$ in the norm of $L^2(g)$.

Theorem 4.18 provides an analogue of Theorem 4.1 [14], for the degenerate rank case. Next we provide an analogue of a theorem due to Masani [6], when $\{X_n\}$ is of degenerate rank. For the time being we assume that f satisfies the following conditions:

(4.20) $$f \in L^{\infty}_{q\times q}, \qquad f^{\#} \in L^{1}_{q\times q}.$$

Therefore, it follows from (4.6) that

(4.21) $$\Phi^{\#} = G^{\#1/2}\Omega^{-1} \in L^2_{q\times q}.$$

Since $\Phi^{\#} = G^{\#1/2}\Omega^{-1} \in L^2_{q\times q}$, from the Riesz-Fischer theorem and the boundedness of f, we conclude that the Fourier series of $\Phi^{\#} = G^{\#1/2}\Omega^{-1}$ converges to $\Phi^{\#} = G^{\#1/2}\Omega^{-1}$ in the norm of $L^2(f)$. However, this does not necessarily imply that the Fourier (Taylor) series of $\Omega^{-1}$ converges to $\Omega^{-1}$ in the norm of $L^2(f)$

(since $G^{\#}$ is not of full rank). The latter convergence is what we need to establish the convergence of the autoregressive representation of $\{X_n\}$ or $\hat{X}_{n+\nu}$, (cf. Corollary 4.16). Thus we need an additional condition so that (4.21) implies that

$$\Omega^{-1} \in L^2_{q\times q} . \tag{4.22}$$

It is easy to check that if f has a constant range then (4.21) implies (4.22). Therefore, we have proved the following

4.23 <u>Theorem</u>. Let $\{X_n\}$ be as a purely nondeterministic WSSP with the spectral density matrix f. If f has a constant range and satisfies (4.20) then,

a) $\{X_n\}$ has a mean-convergent autoregressive representation.

b) For $\nu \geq 1$, $\hat{X}_{n+\nu}$ has a mean-convergent autoregressive representation.

Let $\Lambda$ denote the class of densities for which $\rho(X)<1$, and M denote the class of densities satisfying the conditions of Theorem 4.23. One can define a larger class, denoted by $\Lambda \otimes M$, which contains either of the previous two classes:

$$\Lambda \otimes M = \{f:\ f = f_1^{1/2} f_2 f_1^{1/2},\ f_1 \in \Lambda,\ f_2 \in M\}.$$

Let $\{X_n\}$ be a WSSP with density $f \in \Lambda \otimes M$, then by using the method of proof of Theorem 3.2 in [15], one can show that $\{X_n\}$ or $\hat{X}_{n+\nu}$ has a mean-convergent autoregressive representation. Similarly, one can form even larger classes based on $\Lambda \otimes M$, for which the corresponding processes $\{X_n\}$ admit mean-summable autoregressive representations. Here, we do not discuss the details of these ideas, as already they have been studied in [15]. The results of the later part of this section reveal that the problem of autoregressive representation

of a degenerate rank process $\{X_n\}$ can be reduced to that of a full rank process of smaller dimension, and therefore one may utilize the known results of the latter case. These can be found in [10, 14, 15].

## 5. Appendix; Proof of Theorem 3.6.

Throughout this Appendix $A_n$ and $C_n$, $n = 0, \pm 1, \ldots$, stand for $q \times q$ constant matrices.

First, suppose that there exists a positive number $M$ such that

$$||\sum_{n=k}^{\ell} A_n X_n|| \leq M||\sum_{n=-\infty}^{\infty} A_n X_n||, \quad \text{for all } k \geq \ell. \tag{5.1}$$

Let $Y \in H(x)$, then since $H(x)$ is complete we have

$$Y = \lim_{n\to\infty} Y_n ,$$

with

$$Y_n = \sum_{i=k_n}^{m_n} C_i^{(n)} X_i, \quad n=1,2,\ldots,$$

where $k_n$, $m_n$ are integers. Next, we show that for each fixed i, $\{C_i^{(n)} X_i\}$ is a Cauchy sequence in $H(X)$. For this, note that from (5.1) we have

$$||C_i^{(n)} X_i - C_i^{(n')} X_i|| \leq M||Y_n - Y_{n'}||. \tag{5.2}$$

Since $\{Y_n\}$ is convergent and hence a Cauchy sequence in $H(X)$, it follows from (5.2) that $\{C_i^{(n)} X_i\}$ is a Cauchy sequence in $H(X)$. Thus there exists $Z_i \in H(X)$ such that, for each integer i

$$C_i^{(n)} X_i \to Z_i, \text{ in } H(X), \text{ as } n \to \infty.$$

But since $C_i^{(n)} X_i \in \overline{sp}\ \{X_i\}$ and $Z_i \in \overline{sp}\ \{X_i\}$, we have $Z_i = C_i X_i$, for some matrix $C_i$. Therefore,

$$C_i X_i = \lim_{n\to\infty} C_i^{(n)} X_i .$$

To finish the proof of the first part we need to show that

$$\sum_{i=-\infty}^{\infty} C_i X_i \in H(X), \tag{5.3}$$

and

$$Y = \sum_{i=-\infty}^{\infty} C_i X_i \ . \tag{5.4}$$

For this let $\varepsilon > 0$ be given, then we can choose $k_\varepsilon > 0$ such that

$$|| Y_n - Y_{n'} || < \varepsilon, \text{ whenever } n, n' > k_\varepsilon.$$

Now, for any integers $m_0 \leq m_1$, $m_2 \leq m_3$, by triangle inequality and (5.1) we have

$$|| \sum_{i=m_0}^{m_1} (C_i^{(n)} X_i - C_i^{(n')} X_i) + \sum_{i=m_2}^{m_3} (C_i^{(n)} X_i - C_i^{(n')} X_i) || \leq$$

$$\leq || \sum_{i=m_0}^{m_1} (C_i^{(n)} X_i - C_i^{(n')} X_i) || + || \sum_{i=m_2}^{m_3} (C_i^{(n)} X_i - C_i^{(n')} X_i) || < 2M\varepsilon .$$

Letting $n' \to \infty$, we get

$$|| \sum_{i=m_0}^{m_1} (C_i^{(n)} X_i - C_i X_i) + \sum_{i=m_2}^{m_3} (C_i^{(n)} X_i - C_i X_i) || \leq 2M\varepsilon, \tag{5.5}$$

whenever $n > k_\varepsilon$.

By choosing $m_1$ and $m_2$ large enough so that

$$C_i^{(k_\varepsilon + 1)} X_i = 0, \quad \text{for all } i \notin [m_1, m_2],$$

we get from (5.5) that

$$|| \sum_{i=m_0}^{m_1} C_i X_i + \sum_{i=m_2}^{m_3} C_i X_i || \leq 2M\varepsilon .$$

Therefore $\{ \sum_{i=-n}^{n} C_i X_i \}_{n=1}^{\infty}$ is a Cauchy sequence in $H(X)$. Since $H(X)$ is complete this proves (5.3). It remains to show (5.4).

To this end, for given $\varepsilon > 0$, it follows from (5.3) that there exists $N_\varepsilon > 0$ such that, whenever $m_1, m_2 > N_\varepsilon$, then

$$(5.6) \qquad \left\|\sum_{i=m_0}^{m_1} C_i X_i + \sum_{i=m_2}^{m_3} C_i X_i\right\| < \varepsilon, \quad \text{for all } m_0 \le m_1,\ m_2 \le m_3.$$

Also, for $n > k_\varepsilon$ we have

$$(5.7) \qquad \left\|\sum_{i=m_1}^{m_2} C_i^{(n)} X_i - \sum_{i=m_1}^{m_2} C_i X_i\right\| = \lim_{n'\to\infty} \left\|\sum_{i=m_1}^{m_2} C_i^{(n)} X_i - \sum_{i=m_1}^{m_2} C_i^{(n')} X_i\right\|$$

$$\le M \lim_{n'\to\infty} \|Y_n - Y_{n'}\| \le M\varepsilon,$$

where the first equality holds true because $C_i^{(n')} X_i \to C_i X_i$, for each i, and the inequalities are the result of (5.1) and the choice of $k_\varepsilon$. Now, for $n > k_\varepsilon$, we can choose $m_1$ and $m_2$ large enough so that in addition to (5.6) we have

$$Y_n = \sum_{i=m_1}^{m_2} C_i^{(n)} X_i = \sum_{i=-\infty}^{\infty} C_i^{(n)} X_i \;,$$

where $C_i^{(n)}$'s not present in the original representation of $Y_n$ should be interpreted as zero matrix. Then,

$$\left\|Y_n - \sum_{i=-\infty}^{\infty} C_i X_i\right\| = \left\|\sum_{i=-\infty}^{\infty} C_i^{(n)} - \sum_{i=-\infty}^{\infty} C_i X_i\right\| \le$$

$$\left\|\sum_{i\notin[m_1,m_2]} C_i^{(n)} X_i\right\| + \left\|\sum_{i=m_1}^{m_2} (C_i^{(n)} X_i - C_i X_i)\right\| +$$

$$\left\|\sum_{i\notin[m_1,m_2]} C_i X_i\right\| \le (M+1)\varepsilon.$$

Because the first term on the right hand side of the inequality is zero (recall how $m_1$ and $m_2$ are chosen), the second term is less than $M\varepsilon$ by (5.7), and the third term is less than $\varepsilon$ by the choice of $m_1$ and $m_2$ in (5.5). And this establishes (5.4).

Finally, to show the uniqueness of a representation of Y in the sense of Definition 3.2(b), suppose that

$$Y = \sum_{i=-\infty}^{\infty} C_i X_i = \sum_{i=-\infty}^{\infty} C_i' X_i .$$

We have from (5.1) that

$$|| C_i X_i - C_i' X_i || \leq M||Y - Y|| = 0,$$

which implies $C_i X_i = C_i' X_i$, for all integers i.

To prove the other part of the theorem, assume that $\{X_n\}$ is a generalized Schauder basis and define the space S by

$$S = \{\{C_i X_i\}_{i=-\infty}^{+\infty} ; \sum_{i=-\infty}^{\infty} C_i X_i \varepsilon H(X)\}.$$

It is clear that S is a linear space and the following defines a norm on S

$$||\{C_i X_i\}_{i=-\infty}^{\infty}||_S = \sup_{m_1 \leq m_2} || \sum_{i=m_1}^{m_2} C_i X_i ||_{H(X)} .$$

Now, by using the ideas in the first part of the proof of this theorem, one can show that S with the norm defined above is a Banach space. Consider the operator $T:S \to H(X)$ defined by

$$T(\{C_i X_i\}_{i=-\infty}^{+\infty}) = \sum_{i=-\infty}^{+\infty} C_i X_i .$$

By using the two defining properties of a generalized Schauder basis it can be shown that T is a one-to-on and onto operator. Furthermore, T is bounded, because

$$||T(\{C_i X_i\})||_{H(X)} = || \sum_{i=-\infty}^{\infty} C_i X_i ||_{H(X)} = \lim_{n\to\infty} || \sum_{i=-n}^{n} C_i X_i ||_{H(X)} \leq ||\{C_i X_i\}||_S .$$

Thus by the open mapping theorem $T^{-1}$ is bounded. By choosing $M = ||T^{-1}||$ we get

$$||\{C_i X_i\}||_S \le M||\sum_{i=-\infty}^{\infty} C_i X_i||_{H(X)},$$

or equivalently

$$||\sum_{i=m}^{n} C_i X_i|| \le M||\sum_{i=-\infty}^{\infty} C_i X_i||,$$ for all $m \le n$, which is the same as (5.1). Q.E.D.

# REFERENCES


[1] Bloom, S. (1981). Weighted norm inequalities for vector-valued functions. Ph.D. Thesis. Department of Mathematics, St. Louis, Missouri.

[2] Bloomfield, P. (1985). On series representation for linear predictors. Ann. of Probab., 13, 236-253.

[3] Helson, H. and Szegö, G. (1960). A problem in prediction theory. Ann. Mat. Pura Appl. 51, 107-138.

[4] Lacey, H.E. (1974). The isometric theory of classical Banach spaces. Springer-Verlag, New York.

[5] Makagon, A. and Weron, A. (1976). Wold-Cramér Concordance Theorems for interpolation of q-variate stationary processes over locally compact abelian groups. Journal of Multivariate Analysis, 6, 123-137.

[6] Masani, P. (1960). The prediction theory of multivariate stochastic processes, III. Acta Math. 104, 141-162.

[7] Masani, P. (1966). Recent trends in multivariate prediction theory. Multivariate Analysis (ed. P.R. Krishnaiah) 351-382. Academic Press, New York.

[8] Masani, P. (1981). Commentary on the prediction-theoretic papers. In Norbert Wiener: Collected Works, Volume III. P. Masani (ed.), The MIT Press.

[9] Matveev, R.F. (1961). On multidimensional regular stationary processes. Theory of probability Appl., 6, 149-165.

[10] Miamee, A.G. (1984). On the angle between past and future and prediction theory of stationary stochastic processes. Preprint.

[11] Miamee, A.G. (1985). On the prediction of non-full-rank multivariate stationary stochastic processes. Technical Report #96, Center for Stochastic Processes, Department of Statistics, University of North Carolina at Chapel Hill.

[12] Miamee, A.G. and Salehi, H. (1979). On the bilateral prediction error matrix of a multivariate stationary stochastic process. SIAM J. Math Anal. 10, 247-252.

[13] Pourahmadi, M. (1984). The Helson-Sarason-Szegö Theorem and the Abel summabilityof the series for the predictor. Proc. of the Amer. Math. Soc. 105, 306-308.

[14] Pourahmadi, M. (1985). A matricial extension of the Helson-Szegö theorem and its application in multivariate prediction. J. of Multivariate Analysis, 16.

[15] Pourahmadi, M. (1985). Infinite order autoregressive representation of multivariate stationary stochastic processes. submitted.

[16] Pousson, H.R. (1968). Systems of Toeplitz operator on $H^2$, II. Trans. Amer. Math. Soc. 105, 32-42.

[17] Rosanov, Iu. A. (1967). Stationary random processes. Holden Day, San Francisco.

[18] Wiener, N. and Masani, P. (1958). The prediction theory of multivariate stochastic processes, II. Acta Math. 99, 93-137.

# END

# FILMED

10-85

# DTIC